AWARD NUMBER: W81XWH-12-1-0330

TITLE: Salivary Proteomic and MicroRNA Biomarkers Development for Lung Cancer Detection

PRINCIPAL INVESTIGATOR: David T. Wong, DMD, DMSc

CONTRACTING ORGANIZATION: University of California, Los Angeles
Los Angeles, CA 90095

REPORT DATE: October 2016

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
October 31, 2016	Final	1Aug2012 - 31Jul2016

4. TITLE AND SUBTITLE
Salivary Proteomic and micro-RNA Biomarkers Development for Lung Cancer Detection

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W92XWH-12-1-0330

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
David T. Wong, DMD, DMSc

Email: dtww@ucla.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles
11000 Kinross Avenue, Suite 211
Los Angeles, CA 90095-2000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5014

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This is a lung cancer biomarker development project to test the hypothesis that there are discriminatory biomarkers in saliva that can detect lung cancer with the aim to reduce the number unnecessary diagnostic workups (bronchoscopy) in patients with suspicious chest symptoms. The major goal is to perform a properly powered biomarker discovery and validation of salivary miRNA and proteomic biomarkers for detection of lung cancer based on PRoBE design principles (prospective-specimen-collection and retrospective-blinded-evaluation). The outcome of this project will be a panel of validated non-invasive saliva-based biomarkers for detection of lung cancer.

15. SUBJECT TERMS

Lung cancer, early detection, saliva, biomarkers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U		U	UU	23	19b. TELEPHONE NUMBER *(Include area code)*

Table of Contents

1. **INTRODUCTION:** Narrative that briefly (one paragraph) describes the subject, purpose and scope of the research.

This is a lung cancer biomarker development project to test the hypothesis that there are discriminatory biomarkers in saliva that can detect lung cancer with the aim to reduce the number unnecessary diagnostic workups (bronchoscopy) in patients with suspicious chest symptoms. The major goal is to perform a properly powered biomarker discovery and validation of salivary miRNA and proteomic biomarkers for detection of lung cancer based on PRoBE design principles (prospective-specimen-collection and retrospective-blinded-evaluation). The outcome of this project will be a panel of validated non-invasive saliva-based biomarkers for detection of lung cancer.

2. **KEYWORDS:** Provide a brief list of keywords (limit to 20 words).
Lung cancer, Early detection, Saliva, Biomarkers

3. **OVERALL PROJECT SUMMARY:** Summarize the progress during appropriate reporting period (single annual or comprehensive final). This section of the report shall be in direct alignment with respect to each task outlined in the approved SOW in a summary of Current Objectives, and a summary of Results, Progress and Accomplishments with Discussion. Key methodology used during the reporting period, including a description of any changes to originally proposed methods, shall be summarized. Data supporting research conclusions, in the form of figures and/or tables, shall be embedded in the text, appended, or referenced to appended manuscripts. Actual or anticipated problems or delays and actions or plans to resolve them shall be included. Additionally, any changes in approach and reasons for these changes shall be reported. **Any change that is substantially different from the original approved SOW (e.g., new or modified tasks, objectives, experiments, etc.) requires review by the Grants Officer's Representative and final approval by USAMRAA Grants Officer through an award modification prior to initiating any changes.**

The first year of this lung cancer biomarker development project was spent in the obtainment of regulatory (IRB) approvals from the two performance sites of the project, University of California Los Angeles and the Greater Los Angeles VA (GLA-VA), as well as with the Human Research Protection Office (HRPO) at the US Army Medical Research and Materiel Command (USAMRMC). These lengthy regulatory procedures caused a year of setbacks delaying the initiation our translational research study to develop salivary biomarkers for lung cancer detection. We have since obtained full approval of the informed consent changes and the HRPO of USAMRMC have approved the use of human subjects of this lung cancer biomarker development study. On November 15, 2013 we obtained approval from Dr. Sheilah Rowe, the Scientific Officer, that our project was delayed for one year and consider the need of an extension of the performance period.

The second year of the project was focused on Aim 1 to continue accrual of lung cancer and control subjects. We also began the salivary biomarker discovery phase, Aim 2 of the project. As reported in the second year progress report, significant efforts were made to integrate the emerging technology of RNA-Seq in saliva to discovery miRNA in saliva of lung cancer patients. The efforts to optimize the RNA-seq technology to saliva for extracellular RNA discovery will allow us to obtain un-parallel detailed information of known and novel miRNA in saliva that can be developed for lung cancer non-invasive biomarkers. 30 lung cancer and 30 non-lung cancer control saliva RNA-libraries were constructed for biomarker discovery. A high

impact paper was published based on these novel and impactful efforts to decipher the salivary extracellular transcriptome and the article was featured in a special issue of Clinical Chemistry titled “Molecular Diagnostics: A Revolution in Progress” (1). Support by the CDMRP/DoD was cited.

Aim 1: Accrual of Lung Cancer and Control Subjects- Based on PRoBE Design

Milestone 1: *Accrual of 1560 saliva samples from patients with suspicious chest symptoms. Based on current practice, we anticipate 624 lung cancers and 936 are cancer free patients at the Greater Los Angeles VA hospital (GLA-VA) procured based on the PRoBE-study design.*

As of August 28, 2015 we have screened 2486 patients with chest symptoms at the GLA-VA (159% of the targeted enrollment of 1560). Of these 215 subjects were endoscoped and 93 were confirmed with diagnosed of lung cancer. Our original study design anticipated 624 lung cancer cases by the end of year 01 with nodular sizes on CT > 1cm. The lung cancer yield turned out to be 32 cases with nodular size >1 cm. This is lower than anticipated and necessitated us to modify the study design for the biomarker discovery Aim.

We used 30 lung cancer and 30 non-cancer saliva samples for the biomarker discovery. In addition to the cancer status of these lesions, we have also correlated the tumor size of these lesion based on their CT data. This inclusion is of clinical relevance and impact since the ability to develop salivary biomarkers that can predict cancer from non-cancer patients will be clinically impactful. By examining the plot of sample size against the proportion of genes exceeding the power threshold, we estimated that the sample size of 30 per group (cases and control) will prove statistical power of at least 99% for 98% of the genes whose true effects exceed a fold-change of 2. Saliva from 30 lung cancer patients and 30 matched controls were used for biomarker discovery. Controls were matched for gender, age, smoking history and ethnicity. This matching ensured a distributional match on potential confounders.

Aim 2: Salivary extracellular RNA (exRNA) Biomarkers Discovery, Statistical and Systems Approaches to Candidate Biomarkers Selection

Milestone 2: Optimized salivary biomarker discovery technologies and a systems approach will be used to *identify candidate exRNA salivary biomarkers for lung cancer detection in a discovery cohort of 30 cases and 30 controls randomly selected from Aim 1. Salivary biomarker optimized data mining approach will be used to identify to candidate markers.*

RNA-Seq is emerging technology to obtain the most detailed information of RNA in a biological sample. While we originally proposed to use the Taqman MicroRNA Array Card for saliva miRNA discovery, the significant advantages to use RNA-Seq for saliva miRNA discovery for known and novel miRNA is compelling. We published the first RNA-Seq study on salivary RNA using the SOLID™ system (2). In this project, we used Illumina sequencing systems. We have generated data that support the quality, reproducibility and feasibility of our approach. We have compared multiple library generation methods, constructed different types of libraries to capture the whole spectrum of exRNA in saliva, evaluated the reproducibility of our methods, and obtained a preliminary landscape of relative and absolute concentration of various types of exRNAs in saliva (1).

A number of RNA-Seq library construction methods have been developed in the literature. We have evaluated the performance of alternative methods, we used multiple commercially available kits (NEB, Illumina, Clonetech and NuGen) targeting different types of RNA. A typical bioanalyzer profile of saliva exRNA is shown in Fig. 1. For each library, 500ng of total RNA was used as input. Importantly, predefined amount of synthetic spike-in RNAs were added into each RNA sample equivalently, which will serve as internal standards to evaluate library efficiency, reproducibility, to normalize data across different samples, and to calculate absolute RNA abundance. The synthetic RNAs were purchased from Exiqon and Life Technologies for small RNA-Seq and regular RNA-Seq, respectively. The synthetic RNA pool consists of many distinct RNA species (>40 for small RNA) to ensure abundance and sequence diversity. Since it is known that RNA from saliva is partially degraded with size between 20 and 200nt, we modified the library generation methods to exclude polyA selection and include a size-selection step favoring RNAs below 200nt. Depletion of ribosomal RNA was not carried out since it is known that saliva has relatively less rRNA compared to cellular RNA. Note that although the regular RNA-Seq spike-in RNAs were polyadenylated, the random priming method used in regular RNA-Seq still allows their usage as reference standards. Using these optimized steps for salivary exRNA, we constructed RNA-libraries of the 60 saliva samples (30 lung cancer, 30 controls). All samples were randomized to minimize batched effect for RNA-Sequencing.

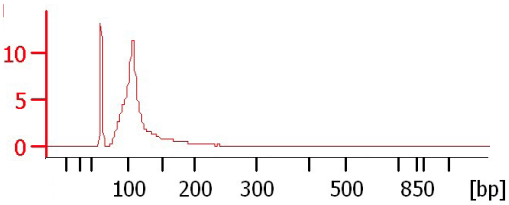

Figure 1. A typical bioanalyzer trace of total RNA from saliva isolated by Trizol LS.

All libraries were sequenced using Illumina HiSeq 2500 sequencers at the UCLA core facility. A total of 30-50 million single-end (50nt) reads were obtained for each library. We developed a customized bioinformatic pipelines to identify different types of non-coding RNAs (ncRNA) present in these data sets originated from human, microbiome, plants or any other species (Fig. 2). Small RNA-Seq data were analyzed for miRNAs and other ncRNAs. Although the small RNA libraries capitalize on the fact that canonical miRNAs have a 5'-phosphate and 3'-OH, other ncRNAs may be identified if their processing steps also lead to such footprints. Since the RNA-Seq libraries used random priming to generate cDNAs, they can theoretically capture all different types of RNAs in the selected size range (20-200nt) with adequate abundance. We have analyzed whether the RNA-Seq data sets may capture all long, small and circular ncRNAs. Mapping uniqueness was required for reads mapped to spike-in RNAs, known genes, lncRNAs and circular RNAs, but not for reads mapped to microbiome or 16S. Small RNA reads were not required to be unique either since small RNAs (miRNAs, piRNAs, etc.) may have multiple copies or similar family members in the human genome. All libraries yielded high quality reads, with an average of ~50% reads mapped to 16S and microbiome. To evaluate potential contamination by cellular RNA in our samples, we examined a number of genes (e.g., ESRP1/2, OVOL1/2, HBA1, APOC1 etc.) that are known to be highly specific to epithelial cells

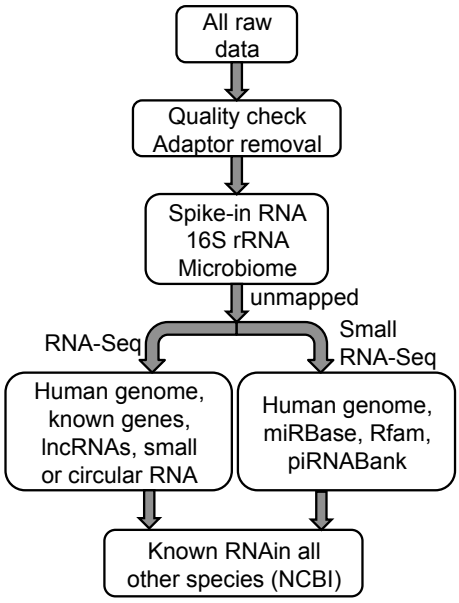

Figure 2. RNA-Seq and small RNA-Seq data analysis. Read mapping was carried out using Bowtie2 allowing up to 1 mismatch in the adaptor-trimmed reads.

or leukocytes, the major types of cells in saliva. Most of these genes are not expressed based on the RNA-Seq data, supporting the effectiveness of our saliva SOP in removing cells.

Sequencing data was preprocessed to generate counts for small exRNA species. Different normalization methods (raw counts, RPM counts, Quantile normalization and DESeq normalization) were used to generate analytic datasets. We filtered the small exRNA species by requiring counts ≥ 5 in at least 20% of samples for inclusion in the statistical analysis. The Wilcoxon rank sum test and the Negative Binomial DE-Seq test were used to compare cancer vs. non-cancer for each species. Initial small exRNA candidates were selected based statistical significance and additional criteria of biological relevance. The first set of candidates was assayed with ddPCR to evaluate optimal data analysis strategies for the candidate selection process. We have found that Quantile normalization of the sequencing data did not correlate well with ddPCR concentrations. Normalizing the ddPCR targets to reference gene also negatively affected the concordance between techniques. We observed correlations ~(0.70-0.90) between untransformed ddPCR concentrations and number of RNA-seq reads (either raw or RPM).

Table 1: Salivary miRNA candidates for lung cancer detection

miRNA	Control % Read Count >=10	Cancer % Read Count >=10	control.median	control.Q1	control.Q3	cancer.median	cancer.Q1	cancer.Q3	P.Wilcox.RC	P.Wilcox.RPM	FC.DESeq	P.DESeq	AdjP.DESeq	FC.edgeR	P.edgeR
miR_4488	36	45	7	1	20	8.5	1.25	26.75	0.723107	0.805098	14.49004	1	1	11.12099	3.32E-05
miR_3656	12	18	1	0	3	1	0	7.25	0.646254	0.568941	29.03614	0.95073	1	10.47615	5.43E-05
miR_10b_5p	16	45	2	1	6	6.5	3	21.75	0.013258	0.022696	2.954894	0.010764	0.194769	3.039285	0.000178
miR_381_3p	44	50	7	3	14	9	4.25	17	0.500384	0.600546	3.099706	0.374787	1	3.261079	0.000328
miR_210	92	77	56	26	101	63.5	26.5	123	0.88134	0.364826	3.796675	0.838827	1	2.419881	0.007018
miR_664a_5p	48	68	9	3	20	18	8.25	29.25	0.223978	0.501726	4.130597	1	1	1.553393	0.015215
miR_1307_3p	96	95	126	66	279	149	54.75	258.25	0.83948	0.546895	0.360961	0.611561	1	0.648972	0.016097
miR_342_5p	64	68	14	8	61	22	7.25	34.25	0.948974	0.272156	0.156915	0.699597	1	0.588872	0.020103
miR_27b_3p	96	100	1813	950	4924	3664	1354.5	9091.5	0.164043	0.335288	1.56169	0.094792	0.837695	1.660131	0.023278
miR_451a	88	82	41	23	341	225.5	31.75	738.5	0.286255	0.39139	2.010255	0.709001	1	2.197283	0.023998
miR_125b_2_3p	36	59	6	3	16	13	4.5	24	0.104639	0.175667	1.459217	0.220288	1	1.876335	0.026635
miR_122_5p	0	9	0	0	2	1	0	3.75	0.325835	0.322104	1.906507	0.042304	0.53585	3.787577	0.035394
miR_660_5p	76	86	53	18	137	107.5	21.75	146.75	0.311006	0.359253	1.469822	0.320974	1	1.569357	0.041835
miR_148a_3p	100	100	25175	8172	51507	47217	13060.25	74902	0.265764	0.214503	1.371327	0.31643	1	1.404031	0.044303
miR_183_5p	88	91	201	118	276	311.5	136.25	729.25	0.122133	0.191678	1.45702	0.238376	1	1.577003	0.046841
miR_100_5p	96	91	111	71	235	171.5	56.75	287	1	0.974736	0.454766	0.667452	1	0.611091	0.049424
miR_2115_5p	44	50	6	3	45	10	2.25	20.25	0.797258	0.645531	0.321632	0.667086	1	0.565609	0.049514
miR_184	88	95	62	28	120	65	42.5	105.75	0.716975	0.874174	0.450008	0.801576	1	0.521706	0.049714

Based on these findings we revised the rationale for miRNA candidates' selection for verification. Table 1 showed the identification of 18 salivary (**miRNA 9, miR 4488, miR 3656, miR 10b 5p, miR 381 3p, miR 210, miR 664a 5p, miR 1307 3p, miR 342 5p, miR 27b 3p, miR 451a, miR 125b 2 3p, miR 122 5p, miR 660 5p, miR 148a 3p, miR 183 5p, miR 100 5p, miR 2115 5p, miR 184**) that are significantly different between the lung cancer and non-lung cancer controls. These 18 miRNAs are being validated in Aim 3.

Aim 3: Salivary Biomarker Validation, Detection Panel Development & Validation

Milestone 3: This aim is to perform validation of the short list of candidate salivary biomarkers *from Aim 2, develop a biomarker panel, and clinically test the panel of salivary biomarkers for lung cancer detection. Phase 1 is individual biomarker validation and panel building using 30 cases and 30 controls. Phase 2 is to validate the fixed panel for its sensitivity and specificity for lung cancer detection in an independent test set of 30 cases and 30 controls.*

Validation of salivary lung cancer biomarkers: The PRoBE-designed cohort in Aim 3 provides important opportunity to validate each candidate salivary miRNA marker, to fine tune parameters for the logistic regression model, and to fix a final biomarker panel for lung cancer detection. Phase 1 is individual biomarker validation and panel building using 30 cases and 30 controls. For each candidate biomarker, we plotted pAUC(0.05), i.e. partial area under ROC curve at false positive rate 5% (specificity 95%) curve and construct its 95% confidence interval

using method of Pepe at al. (3). pAUC as the selection criterion has the advantage of focusing highly specific biomarkers that explain sub-class of tumor. Individual biomarker was considered as validated for its association with lung cancer if the low bound of the 95% confidence interval for pAUC(0.05) is above 0.05. If a biomarker is non-informative for disease, its expected sensitivity will be 5% at threshold corresponding to 95% specificity. Ten of the 18 candidate miRNA markers were individually validated based on these criteria: **miRNA 9, miR 381 3p, miR 210, miR 342 5p, miR 27b 3p, miR 451a, miR 660 5p, miR 148a 3p, miR 2115 5p and miR 184**. The performance of the ten validated salivary miRNA biomarkers are shown in Table 2:

Table 2: Performance of Validated Individual Salivary RNA

Salivary miRNA	P-value	AUC (95% CI)
miRNA 9	0.02	0.71 (0.59-0.82)
miRNA 381 3p	0.01	0.74 (0.63-0.85)
miRNA 210	0.01	0.72 0.61-0.83)
miRNA 27b 3p	0.04	0.66 (0.55-0.78)
miRNA 342 5p	0.03	0.76 (0.65-0.86)
miRNA 451a	0.01	0.73 (0.62-0.84)
miRNA 660 5p	0.02	0.59 (0.46-0.71)
miRNA 184a 3p	0.01	0.82 (0.72-0.88)
miRNA 2115 5p	0.04	0.85 (0.78-0.89)
miRNA 184	0.02	0.60 (0.74-0.72)

A multivariate logistic regression using all validated biomarkers was developed. In logistic regression modeling, 10-fold cross-validation was used to assess performance and to minimize overfitting. The most discriminatory panel consisted of 4 miRNAs: **miRNA 9, miR 210, miR 2115 5p and miR 184 with an AUC of 0.89 (sensitivity 0.91 and specificity of 0.76).** This model is now fixed for will be validated in Phase 2. We fully expect we will successfully achieve such performance and complete the panel validation in the next 6 months.

Table 3. Performance of Salivary Biomarkers Combination

Salivary RNA	AUC (95% CI)	Sensitivity	Specificity
miRNA 9 + miRNA 27b 3p + miRNA 2115 5p	0.78 (0.68-0.82)	0.87	0.71
miRNA 381 3p + miRNA 184a 3p + miRNA 2115 5p	0.82 (0.76-0.83)	0.89	0.70
miRNA 2115 5p + miRNA 451a + miRNA 210	0.77 (0.74 to 0.80)	0.86	0.52
miRNA 9 + miR 210 + miR 2115 5p + miR 184	0.89 (0.84- 0.92)	0.91	0.76

4. **KEY RESEARCH ACCOMPLISHMENTS:** Bulleted list of key research accomplishments emanating from this research. Project milestones, such as simply completing proposed experiments, are not acceptable as key research accomplishments. Key research accomplishments are those that have contributed to the major goals and objectives and that have potential impact on the research field.

• Accrual of 2486 patients with chest symptoms (159% of the targeted enrollment of 1560)

• Biomarker discovery cohort of 30 lung cancer patients with lung nodules on CT> 1cm and 30 non-lung cancer matched controls fully adhering to prospective-specimen-collection and retrospective-blinded-evaluation (PRoBE) design

• RNA library construction of the 30 lung cancer and 30 non lung cancer controls

• RNA-Sequencing of 30 lung cancer and 30 non lung cancer controls

• Data analysis and select 18 salivary miRNAs that are significantly altered in lung cancer from non lung cancer controls

• Validation of 10 of the 18 candidate salivary miRNA for lung cancer detection

5. CONCLUSION: Summarize the importance and/or implications with respect to medical and /or military significance of the completed research including distinctive contributions, innovations, or changes in practice or behavior that has come about as a result of the project. A brief description of future plans to accomplish the goals and objectives shall also be included.

During three years of the project, scientific progress has been sound. Targeted enrollment has been attained despite the lung cancer cases fulfilling the inclusion criteria was less than expected. We have successfully performed, using optimized salivary RNA library construction and RNA-sequencing technologies for the biomarker discovery cohort of 30/30. Data analysis of the RNA-seq data revealed 18 salivary miRNAs are significantly altered in lung cancer from non-lung cancer subjects. Ten of the 18 candidate salivary miRNAs have been independently validated. Future plan is to validate the locked down panel of 4 salivary miRNA (**miRNA 9, miR 210, miR 2115 5p and miR 184)** in an independent cohort.

6. PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS:

a. List all manuscripts submitted for publication during the period covered by this report resulting from this project. Include those in the categories of lay press, peer-reviewed scientific journals, invited articles, and abstracts. Each entry shall include the author(s), article title, journal name, book title, editors(s), publisher, volume number, page number(s), date, DOI, PMID, and/or ISBN.

(1) Lay Press:
(2) Peer-Reviewed Scientific Journals:
(3) Invited Articles:
(4) Abstracts:

Bahn JH, Zhang Q, Li F, Chan TM, Lin X, Kim Y, Wong DT, Xiao X. The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. Clin Chem. 2015;61(1):221-30. PubMed PMID: 25376581.
http://www.ncbi.nlm.nih.gov/pubmed/25376581

b. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.

2015 Speaker, “EFIRM-Liquid Biopsy (eLB)”, University of Utah Medical Center, Salt Lake City Utah. September 2, 2015

2015 Keynote speaker, “Saliva-Based Point-of-Care Diagnostics & Global Health”. Point-of-Care Diagnostics & Global Health 2015 World Congress, San Diego CA, September 29-30, 2015

2015 Invited speaker, “EFIRM-Liquid Biopsy (eLB)”, Pathology Grand Round, Department of Pathology, Harbor-UCLA Medical Center. October 9, 2015

2015 Speaker, “EFIRM-Liquid Biopsy (eLB)”. National Cheng Kung University Hospital. Tainan, Taiwan. October 20. 2015.

2015 Speaker, “EFIRM-Liquid Biopsy (eLB)”, IBC Drug Discovery: Advances in Biomarkers and Breakthrough Therapeutics Conference, San Francisco CA November 5-6, 2015

2015 Keynote speaker, “Saliva-Based Detection of Actionable EGFR Mutations in Lung Cancer Patients”, BIOFLUID Biopsies, Boston MA November 16-17, 2015

2015 Dean’s Seminar speaker, “Saliva-Liquid Biopsy”, Oregon Health Science University. November 30, 2015

2015 Keynote Speaker, “EFIRM Liquid Biopsy”, Second North America Saliva Symposium, University of Washington, Seattle WA. December 4, 2015.

2016 AAAS Symposium. Oral Cancer: Epidemiology, Mechanisms and Early Detection. Washington DC. February 11-15 2016.

2016 Speaker, “EFIRM-Liquid Biopsy (eLB)”, SELECTBIO “ Circulating Biomarkers World Congress 2016”, Madrid Spain, March 19, 2016.

2016 Keynote speaker & Chair, “Saliva-Based Detection of Actionable EGFR Mutations in Lung Cancer Patients”, Exosomes & Liquid Biopsy, Taipei Taiwan. April 7-8, 2016

2016 Speaker, “EFIRM Liquid Biopsy (eLB) to detection actionable mutations in NSCLC”. ASCO, Chicago. June 4, 2016.

2016 Speaker, “Saliva Exosome Liquid Biopsy”, SELECTBIO, Cambridge, UK. July 9-12 2016

2016 Keynote Speaker, “Saliva Diagnostics: Mapping for the Future, Detecting Diseases Earlier for Better Overall Health Outcomes”, National Dental Association Annual Convention. Atlanta GA. July 25 2016

2016 Symposium speaker, “ Personalized, Genotype-Directed Treatment and Liquid Biopsy for Non Small Cell Lung Cancer (NSCLC)”, American Association for Clinical Chemistry (AACC), 68th Annual Scientific Meeting & Clinical Lab Expo, Philadelphia PA. August 3, 2106.

2016 Speaker, “Saliva Biomarkers for Sjögren’s Syndrome Detection”, UCLA Department of Rheumatology, Grand Round. Los Angeles, CA. August 17, 2016

2016 Discussion Leader: Excellular RNA. Gordon Research Conference, Newry Maine, August 21-26, 2016.

2016 Keynote Speaker, “Saliva Liquid Biopsy”, European Association of Oral Medicine, 13th Biannual Congress. Torino Italy. September 16, 2016

2016 Invited Speaker, “Saliva as A Biomarker Source for Point-of-Care Applications”, Biomarkers and Targeted Therapeutics in Sjögren’s. Oklahoma City, Oklahoma. September 19-22, 2016

2016 Keynote Speaker, “Saliva Point-of-Care Diagnostics”, SELECTBIO, San Diego. Congress” San Diego CA September 27, 2016

2016 Keynote Speaker, “EFIRM Liquid Biopsy”, SELECTBIO, San Diego. Congress” San Diego CA September 28, 2016

2016 Keynote Speaker, “EFIRM Liquid Biopsy”, SELECTBIO Diagnostics Summit. Congress” Singapore October 12-13, 2016

7. **INVENTIONS, PATENTS AND LICENSES:** List all inventions made and patents and licenses applied for and/or issued. Each entry shall include the inventor(s), invention title, patent application number, filing date, patent number if issued, patent issued date, national, or international.

   Nothing to report.

8. **REPORTABLE OUTCOMES:** Provide a list of reportable outcomes that have resulted from this research. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life. This list may include development of prototypes, computer programs and/or software (such as databases and animal models, etc.) or similar products that may be commercialized.

   Nothing to report.

9. **OTHER ACHIEVEMENTS:** This list may include degrees obtained that are supported by this award, development of cell lines, tissue or serum repositories, funding applied for based on work supported by this award, and employment or research opportunities applied for and/or received based on experience/training supported by this award.

   For each section, 4 through 9, if there is no reportable outcome, state "Nothing to report."

   Nothing to report.

10. **REFERENCES:** List all references pertinent to the report using a standard journal format (i.e., format used in *Science*, *Military Medicine*, etc.).

1. Bahn JH, Zhang Q, Li F, Chan TM, Lin X, Kim Y, Wong DT, Xiao X. The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. Clin Chem. 2015;61(1):221-30. doi: 10.1373/clinchem.2014.230433. PubMed PMID: 25376581; PMCID: 4332885.

2. Spielmann N, Ilsley D, Gu J, Lea K, Brockman J, Heater S, Setterquist R, Wong DT. The human salivary RNA transcriptome revealed by massively parallel sequencing. Clin Chem. 2012;58(9):1314-21. doi: 10.1373/clinchem.2011.176941. PubMed PMID: 22773539.

3. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. Am J Epidemiol. 2008;167(3):362-8. Epub 2007/11/06. doi: kwm305 [pii]10.1093/aje/kwm305. PubMed PMID: 17982157.

11. **APPENDICES:** Attach all appendices that contain information that supplements, clarifies or supports the text. Examples include original copies of journal articles, reprints of manuscripts and abstracts, a curriculum vitae, patent applications, study questionnaires, and surveys, etc.

J. H. Bahn *et al.*, The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. *Clin Chem* **61**, 221-230 (2015).

Papers in Press. Published November 6, 2014 as doi:10.1373/clinchem.2014.230433
The latest version is at http://hwmaint.clinchem.org/cgi/doi/10.1373/clinchem.2014.230433

Clinical Chemistry 61:1
000–000 (2015)

The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva

Jae Hoon Bahn,[1,2†] Qing Zhang,[1,2†] Feng Li,[3] Tak-Ming Chan,[1,2] Xianzhi Lin,[1,2] Yong Kim,[3,4,5] David T.W. Wong,[2,3,4,6,7*] and Xinshu Xiao[1,2,4,6*]

BACKGROUND: Extracellular RNAs (exRNAs) in human body fluids are emerging as effective biomarkers for detection of diseases. Saliva, as the most accessible and noninvasive body fluid, has been shown to harbor exRNA biomarkers for several human diseases. However, the entire spectrum of exRNA from saliva has not been fully characterized.

METHODS: Using high-throughput RNA sequencing (RNA-Seq), we conducted an in-depth bioinformatic analysis of noncoding RNAs (ncRNAs) in human cell-free saliva (CFS) from healthy individuals, with a focus on microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), and circular RNAs (circRNAs).

RESULTS: Our data demonstrated robust reproducibility of miRNA and piRNA profiles across individuals. Furthermore, individual variability of these salivary RNA species was highly similar to those in other body fluids or cellular samples, despite the direct exposure of saliva to environmental impacts. By comparative analysis of >90 RNA-Seq datasets of different origins, we observed that piRNAs were surprisingly abundant in CFS compared with other body fluid or intracellular samples, with expression levels in CFS comparable to those found in embryonic stem cells and skin cells. Conversely, miRNA expression profiles in CFS were highly similar to those in serum and cerebrospinal fluid. Using a customized bioinformatics method, we identified >400 circRNAs in CFS. These data represent the first global characterization and experimental validation of circRNAs in any type of extracellular body fluid.

CONCLUSIONS: Our study provides a comprehensive landscape of ncRNA species in human saliva that will facilitate further biomarker discoveries and lay a foundation for future studies related to ncRNAs in human saliva.

© 2014 American Association for Clinical Chemistry

Human saliva has been used increasingly for biomarker development to enable noninvasive detection of diseases. The term "salivaomics" was coined to highlight the omics constituents in saliva that can be used for biomarker development and personalized medicine *(1)*. Salivary extracellular RNA (exRNA)[8] was discovered 10 years ago; since then, the nature, origin, and characterization of salivary RNA have been actively pursued *(2–8)*. These studies have demonstrated the potential for the use of salivary RNA to detect oral cancer *(2, 9)*, Sjögren syndrome *(3)*, resectable pancreatic cancer *(7)*, lung cancer *(10)*, ovarian cancer *(11)*, and breast cancer *(12)*. Facilitated by next-generation sequencing technologies, a complex compositional profile of salivary RNA molecules has emerged, encompassing mRNAs, microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), etc. *(8, 13, 14)*. However, the entire spectrum of exRNA from saliva has not been fully discovered, thus warranting further comprehensive deciphering and analyses. In addition, there is also increasing interest in understanding the functional aspects of salivary RNAs in oral and systemic biology. Such studies will be facilitated by a detailed delineation of the landscapes of salivary exRNA.

Although RNA sequencing (RNA-Seq) technologies have been applied to study RNA expression in several body fluids *(8, 15–18)*, most of these studies have focused on analyses of miRNAs and mRNAs. In addition, more recently, thousands of exonic circular RNAs (circRNAs) have been discovered in various human cell types, many of which are highly stable, abundant, and evolutionarily conserved *(19)*. Two circRNAs have

[1] Department of Integrative Biology and Physiology, [2] Molecular Biology Institute, [3] School of Dentistry, [4] Jonsson Comprehensive Cancer Center, [5] Broad Stem Cell Research Center, [6] School of Engineering, [7] Department of Head & Neck Surgery/Otolaryngology, UCLA, Los Angeles, CA.
* Address correspondence to this author at: Department of Integrative Biology and Physiology, UCLA, 611 Charles Young Dr, Los Angeles, CA 90095. E-mail gxxiao@ucla.edu.
† J.H. Bahn and Q. Zhang contributed equally to this work.

Received July 16, 2014; accepted October 2, 2014.
Previously published online at DOI: 10.1373/clinchem.2014.230433

[8] Nonstandard abbreviations: exRNA, extracellular RNA; miRNA, microRNA; snoRNA, small nucleolar RNA; RNA-Seq, RNA sequencing; circRNA, circular RNA; ncRNA, noncoding RNA; CFS, cell-free saliva; UCLA, University of California–Los Angeles; piRNA, piwi-interacting RNA; RPM, reads per million mapped reads; ddPCR, droplet digital PCR; ISI, individual specificity index; CSF, cerebral spinal fluid; ES, embryonic stem; lncRNA, long noncoding RNA.

Copyright (C) 2014 by The American Association for Clinical Chemistry

been shown to act as miRNA sponges, thus playing a role in mediating miRNA targeting *(20)*. It is expected that additional functions of circRNAs may be described soon *(19)*. The stable nature of circRNAs makes these moieties intriguing candidates as functional molecules in circulating body fluid.

We performed a comprehensive analysis of extracellular noncoding RNAs (ncRNAs) in cell-free saliva (CFS) by next-generation sequencing. In addition, we carried out a genome-wide analysis of possible existence of circular RNAs in CFS with RNA-Seq. To our best knowledge, this is the first report and validation of the existence of circRNAs in any body fluid. Our findings credential the presence of salivary ncRNA and, importantly, pave ways for further functional, biological, and biomarker discoveries related to ncRNAs in human saliva.

Materials and Methods

SALIVA COLLECTION AND PROCESSING AND RNA ISOLATION

Unstimulated saliva samples were obtained from healthy volunteers in accordance with a protocol approved by the University of California–Los Angeles (UCLA) Institutional Review Board as described previously *(21)*. More details are included in Supplemental Methods, which accompanies the online version of this article at http://www.clinchem.org/content/vol61/issue1.

CONSTRUCTION OF SMALL RNA-SEQ LIBRARIES

We isolated total RNA directly from CFS as described above. Spike-in RNAs (Exiqon) were added to the total RNA samples (1 reaction volume per microgram RNA) before library construction as internal controls. With the spiked total RNA samples, we prepared small RNA-Seq libraries with the NEBNext Small RNA library Prep kit (NEB). The final libraries were purified with 6% PAGE gel.

CONSTRUCTION OF circRNA-SEQ LIBRARIES

To obtain enriched circular RNAs from the total RNA samples, we used 3 U/μg RNase R (Epicentre) to treat the total RNA (from CFS directly) for 20 min at 37 °C. Subsequently, RNA was extracted with acid phenol/chloroform (pH 4.5). We prepared sequencing libraries with the NEBNext Ultra directional RNA library Prep kit (NEB) followed by AMPure XP Beads size selection (Beckman Coulter).

CFS SMALL RNA-SEQ DATA ANALYSIS

Small RNA-Seq reads were first processed to remove adapter sequences and low quality reads. The reads were then aligned to the human genome with Bowtie *(22)* allowing at most 1 mismatch. We parsed the mapping results to identify reads mapped to miRNAs (miRBase, release 19), piwi-interacting RNAs (piRNAs) (piRNABank, November 2013), and other known small RNAs (RFam, version 11.0). We also mapped reads to the Human Oral Microbiome Databases *(23)* to eliminate those that possibly originated from microbial species. For human miRNAs, only uniquely mapped reads were retained. Uniqueness was not required for reads mapped to piRNAs or other small RNAs owing to their repetitive nature and/or presence of multiple copies in the genome. In parallel, reads were also aligned to the spike-in controls, allowing no mismatches. The number of reads mapped to each miRNA was normalized with the spike-in controls and total number of mapped reads in each library.

Detailed bioinformatic methods, exosome isolation, and experimental validation procedures are described in online Supplemental Methods.

Results and Discussion

SMALL RNA SEQUENCING OF CFS

We used our widely adopted protocol to isolate CFS from fresh saliva samples (see online Supplemental Methods) *(9)*. The protocol previously was shown to effectively remove cells as evaluated by exclusion of cellular genomic DNA and cell counting *(9)*. In addition, we examined several genes [e.g., *ESRP1/2* (epithelial splicing regulatory protein 1 and 2),[9] *OVOL1/2* (ovo-like zinc finger 1 and 2), *HBA1* (hemoglobin, α1), *APOC1* (apolipoprotein C-1)] that are known to be highly specific to epithelial cells or leukocytes, the major types of cells in saliva. Most of these genes were not expressed (on the basis of RNA-Seq data collected for a separate study, data not shown), supporting the effectiveness of our CFS protocol in removing cells. We obtained a total of 165 million reads from 8 CFS small RNA sequencing libraries (see online Supplemental Table 1). Total RNA was isolated directly from the CFS, and synthetic small RNAs were added into the total RNA samples before library construction to serve as spike-in controls. As shown in online Supplemental Fig. 1, expression levels of spike-in controls were highly correlated between samples, supporting the technical consistency of our data. After adapter removal, the reads showed a length distribution that peaked at 22 nt (Fig. 1A), consistent with the expected length of miRNAs. Interestingly, we also observed a second peak at 29 nt that may correspond to piRNAs. The most

[9] *ESRP1/2*, epithelial splicing regulatory protein 1 and 2; *OVOL1/2*, ovo-like zinc finger 1 and 2; *HBA1*, hemoglobin, α1; *APOC1*, apolipoprotein C-1; *EIF3E*; eukaryotic translation initiation factor 3, subunit E; *DDOST*; dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit (non-catalytic).

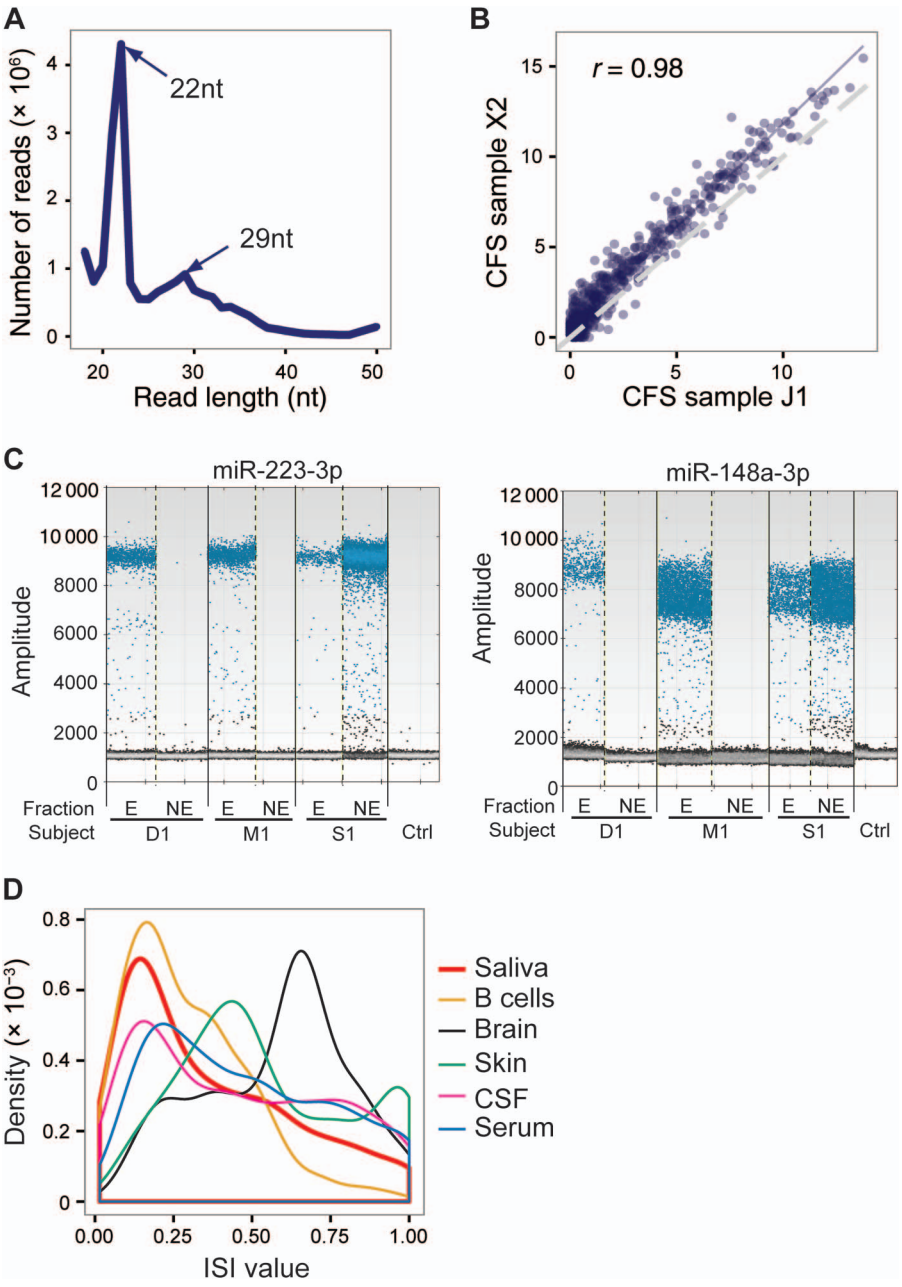

Fig. 1. miRNA expression in CFS.

(A), Example length distribution of a small RNA sequencing library from CFS. Library adapters have been trimmed. The read lengths of the major peak (22 nt) and minor peak (29 nt) are illustrated, which correspond to the known lengths of miRNAs and piRNAs, respectively. (B), Scatter plot of miRNA expression (log2 RPM) across 2 individuals. Pearson correlation coefficient is shown. (C), Experimental validation of miRNA expression in exosome fraction (E) and exosome-free fraction (NE) by use of ddPCR. For each miRNA, the ddPCR fluorescence intensity is shown for E and NE samples in each individual. Negative control (no template) was run together with the actual samples in the same batch of experiments, and fluorescent signal was barely detected. (D), Histogram of ISI for all expressed miRNAs calculated with the 6 independent saliva samples (biological replicates not included). The distributions of ISI values of other public data sets are also shown for comparison.

abundant types of small RNAs in our data included human miRNAs (6.0% of reads on average), piRNAs (7.5% of reads), and snoRNAs (0.02% of reads) (see online Supplemental Table 1). In addition, 58.8% of reads corresponded to microbial RNA sequences, reflecting the enriched presence of microorganisms in saliva *(8)*. Our results suggest that the small RNA sequencing experiment can capture a wide spectrum of noncoding exRNAs in human saliva.

miRNAS ARE ABUNDANT IN HUMAN CFS

In each saliva sample, a total of 127–418 miRNAs were detected, with an expression level of ≥1 reads per million mapped reads (RPM) (see online Supplemental Table 2). The most abundant miRNA was miR-223–3p that had an average expression level of 19 442 RPM across all samples. miRNA expression levels were highly correlated across biological replicates ($r \geq 0.977$) (see online Supplemental Fig. 2), again supporting the reproducibility of our data. Importantly, different human subjects also demonstrated highly correlated miRNA expression levels (Fig. 1B; online Supplemental Fig. 3). For example, among the top 10 highly expressed miRNAs in each sample, 8 were shared by at least 4 of the 6 nonreplicated samples.

Previous reports suggested that the majority of miRNAs in saliva were concentrated in exosomes *(24)*. We tested the presence of miRNAs in exosome and nonexosome fractions of saliva in a subset of samples (Fig. 1C; online Supplemental Fig. 4). Two miRNAs (miR-223–3p and miR-148a-3p) detected in RNA-Seq data of multiple individuals were chosen for this experiment. Plasma/serum miR-223 together with other miRNAs have been shown to be closely associated with the tumorigenesis and metastasis of gastric carcinoma *(25)*, hepatocellular carcinoma or chronic hepatitis *(26)*, sepsis *(27)*, and lung cancer *(28)*. Dysregulated miR-148a has been reported in ovarian cancer *(29)*, liver injury *(30)*, and gastric cancer *(31)*.

To measure expression levels of the 2 miRNAs, exosomes were isolated from fresh human saliva with the conventional differential centrifugation method (see online Supplemental Methods), the effectiveness of which has been well established in our laboratory *(32–34)*. We used the droplet digital PCR (ddPCR) method because it can measure absolute concentration of each miRNA and does not need internal controls. As shown in Fig. 1C, both miRNAs can be detected in all 3 subjects, and they are predominantly present in the exosomal RNA fraction in 2 of 3 subjects. For 1 subject (S1), both miRNAs were detected in both exosomal and nonexosomal fractions. This observation is not likely to be due to failed separation of the exosomal fractions, because piRNAs in this subject showed predominant localization in the exosomal fraction (see below). It should be noted that these data only serve as a qualitative evaluation rather than a quantitative validation of the RNA-Seq data, because the RNA was obtained in very different ways in the 2 types of experiments. In addition, the exosome isolation step introduced relatively large technical variation across samples, which may explain the large interindividual variation observed in miRNA expression level. Nevertheless, our result is consistent with previous findings that miRNAs are mainly localized in exosomes in most individuals. However, as shown in Fig. 1C, it is likely that there also exist vesicle-free ncRNAs in saliva (e.g., for subject S1), which should be further investigated.

VARIATION OF miRNA EXPRESSION ACROSS INDIVIDUALS

As shown in Fig. 1B, miRNA expression values are generally significantly correlated across individuals, as measured by RNA-Seq of CFS total RNA. However, there does exist noticeable variation across different subjects. We asked whether salivary miRNAs are particularly subject to individual variability, given that saliva is readily exposed to and communicative with the external environment that can be highly individual specific. To examine this question, we defined an individual specificity index (ISI) for each miRNA (see online Supplemental Methods). This index has a value between 0 and 1, with larger values representing higher interperson variability. For comparison, we also analyzed several other data sets derived from primary brain tissues *(35)*, skin *(36)*, B cells *(37)*, cerebral spinal fluid (CSF), and serum *(15)*. As shown in Fig. 1D, miRNAs generally demonstrated a wide range of individual variability in all types of samples. Interestingly, the salivary ISI distribution was relatively similar to those of other body fluids (CSF and serum) and intracellular RNA (B cells). Compared with other samples, brain and skin miRNAs showed higher individual variability, possibly reflecting the heterogeneous cell type composition in these samples. Overall, our results suggest that the individual variability observed in salivary miRNAs is at a level similar to those observed in other extracellular and intracellular data sets. Given the remarkable diversity of salivary environment across individuals, this observation supports the effectiveness of our cell removal method in enriching for physiological extracellular RNA rather than environmentally related RNA. Thus, extracellular RNA in CFS can serve as stable biomarkers with individual variability similar to that of other body fluids, with the advantage of being highly accessible and noninvasive.

miRNA EXPRESSION PROFILES OF CFS CLUSTER WITH THOSE OF OTHER BODY FLUIDS

We next conducted a comprehensive comparison of miRNA expression profiles in different body fluids and

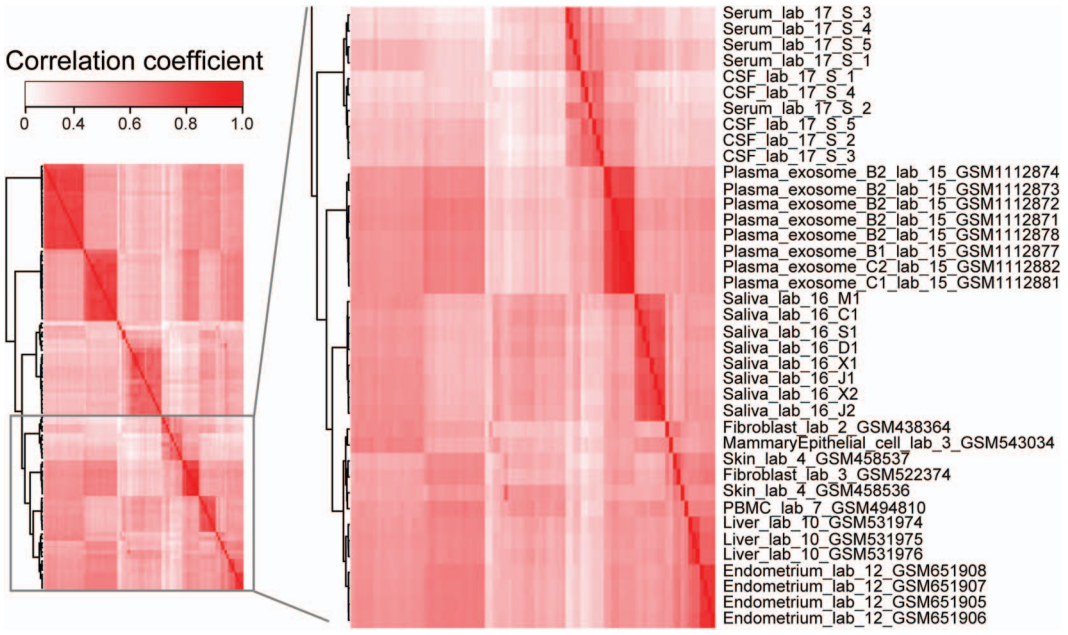

Fig. 2. Comparison of miRNA expression across different cell types and body fluids. Heat map of correlation (Kendall τ) of miRNA expression levels derived from public and in-house small RNA sequencing data sets.

A subset of the samples including our CFS samples is shown here because of space limits, with the entire figure shown as online Supplemental Fig. 3. Hierarchical clustering was applied. Samples were named by their type, the laboratory that generated the data (via an arbitrary numerical ID), and GEO IDs (if available). Raw sequencing data were analyzed in exactly the same way except for data from lab 17, for which expression data of miRNAs in serum and CSF were directly obtained from their publication due to lack of raw data. Saliva_lab_16 represents CFS miRNA data generated in this study. All the data sets derived from extracellular body fluids [CFS, serum, CSF, and plasma (exosome-associated RNA)] clustered together, with relatively smaller distances compared with their distances to intracellular RNA samples.

cell types. A total of 95 small RNA sequencing data sets (including our 8 data sets) were analyzed for miRNA expression with the same method (Fig. 2; online Supplemental Fig. 3; online Supplemental Methods). Raw sequencing reads were used, except for a few data sets in which read counts of miRNAs were directly taken from the original publication owing to lack of raw data (as noted in online Supplemental Fig. 3). Batch effects or technical variations across laboratories may be a significant confounding factor in this type of analysis. We used a data normalization method similar to that adopted in DESeq *(38)* to alleviate batch effects. In addition, the correlation across data sets was evaluated with the Kendall τ method, a rank-based nonparametric correlation analysis (see more details in online Supplemental Methods). In this manner, data comparison is less sensitive to the quantitative values of miRNA expression levels, which may fluctuate due to technical variation.

As shown in online Supplemental Fig. 3, data sets generated from the same tissue or cell type (brain, B cells, or ES cells) by different laboratories clustered together. This observation suggests that batch effects have been adequately reduced, although it may not have been possible to reach a complete elimination. Strikingly, all the data sets derived from extracellular body fluids [CFS, serum, CSF, and plasma (exosome-associated RNA)] clustered together, with relatively smaller distances compared with their distances to intracellular RNA samples. This observation could not have been due to batch effects, since different laboratories generated the data sets of various body fluids. Thus, our analysis suggests that extracellular miRNAs in different body fluids share similar profiles, indicating existence of commonality in the biogenesis of these miRNAs.

Nevertheless, miRNAs in CFS may also have distinct expression patterns that reflect the local cellular environment of the salivary glands and oral mucosa. Epithelial and other cells may release cellular miRNAs into the extracellular space and contribute to the miRNA profile in CFS. Interestingly, among cellular

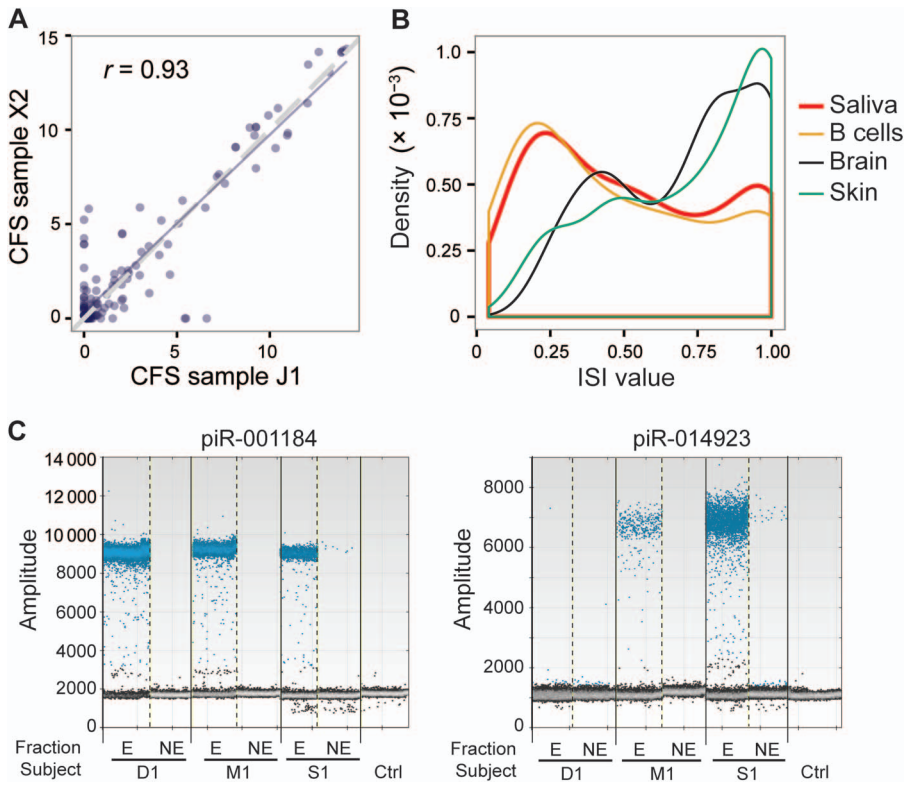

Fig. 3. piRNA expression in CFS.

(A), Scatter plot of piR expression (log2 RPM) across 2 individuals. Pearson correlation coefficient is shown. (B), Histogram of ISI for all expressed piRNAs calculated with the 6 independent saliva samples (biological replicates were not included). The distributions of ISI values of other public data sets are also shown for comparison. (C), Experimental validation of piRNA expression in exosome fraction (E) and exosome-free fraction (NE), similar to Fig. 1C.

data sets that are relatively close to the CFS samples in the clustered heat map (Fig. 2; online Supplemental Fig. 3), several may contain cell types similar or related to the cellular environment of saliva, such as mammary epithelial cells, skin cells, endometrium (with epithelial cells as a major layer), fibroblasts, etc.

piRNAS ARE RELATIVELY ABUNDANT IN HUMAN CFS

As mentioned above, piRNAs constitute another group of small RNAs in CFS with an appreciable peak in the length distribution of sequencing reads (Fig. 1A). A total of 32–109 piRNAs were detected in each saliva sample with at least 1 RPM. Compared with the total number of piRNAs in public databases (23 439 in piRNABank *(39)*), the number of piRNAs detected in our study was small. Nevertheless, some of the piRNAs were expressed at relatively high expression levels in CFS. For example, the most abundant piRNA, piR-018570, had an average expression level of 32 296 RPM across all samples (see online Supplemental Table 3). The expression levels of piRNAs across different CFS samples were highly correlated (Fig. 3A; online Supplemental Fig. 5), although not as strongly as that for miRNAs, possibly because of the relatively small number of piRNAs at high expression levels.

The ISI values of piRNAs in CFS were overall similar to those in B cells (Fig. 3B) but lower than in brain or skin tissues. This observation is similar to that for miRNAs (Fig. 1D). However, the ISI values of piRNAs in CFS were generally larger than those of CFS miRNAs, indicating a higher degree of individual variability. This result may be explained by the possible diversity of cellular origins of CFS piRNAs, which is discussed below.

To test the presence of piRNAs in exosomes, 2 highly expressed piRNAs (piR-001184 and piR-014923) were chosen for ddPCR analysis (Fig. 3C). Both piRNAs were identified in at least 2 subjects, which demonstrated predominant localization in exosomal RNA fractions. Interestingly, in contrast to the

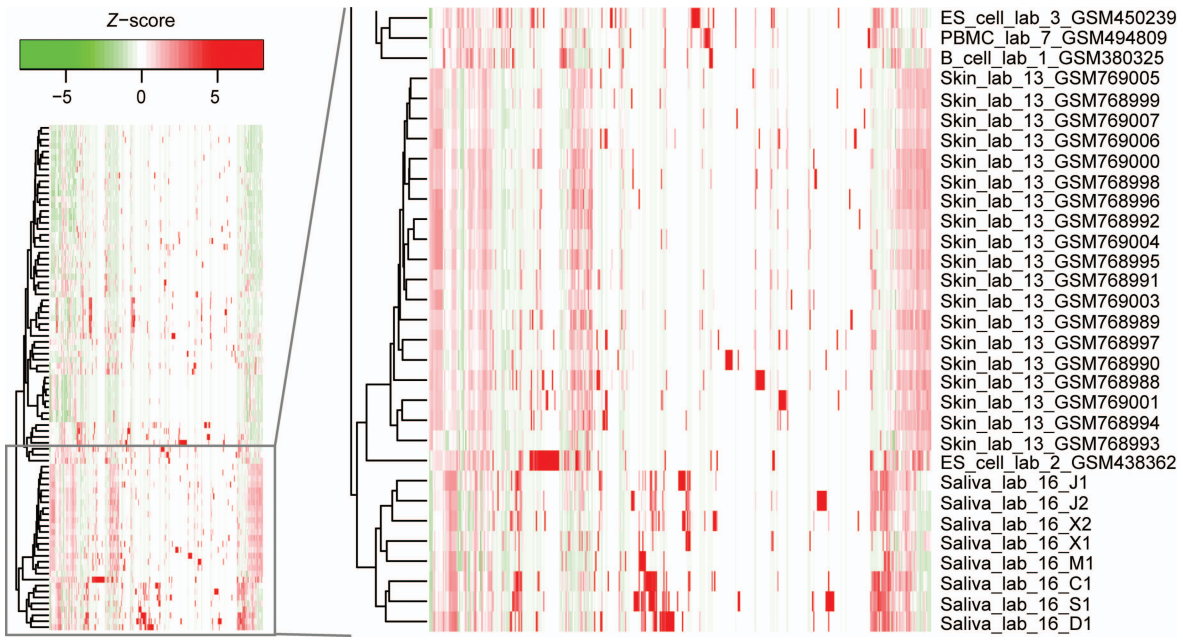

Fig. 4. Comparison of piRNA expression across different cell types and body fluids. Heat map of piRNA expression levels derived from public and in-house small RNA sequencing data sets.

Z-scores of expression levels were calculated for each piRNA. A subset of the samples including our CFS samples is shown here because of space limits, with the entire figure shown as online Supplemental Fig. 5. Hierarchical clustering was applied. Samples were named similarly as in Fig. 2. CFS, ES cells, and skin cells were among those having the highest expression levels of piRNAs.

ddPCR results of miRNAs in subject S1, both piRNAs were mainly detected in exosomes of this subject. Overall, our data suggest that the 2 piRNAs in this experiment had a predominant localization in exosomes in all subjects with detectable signals.

COMPARISON OF piRNA EXPRESSION PROFILES OF CFS AND OTHER SAMPLES

As for miRNAs, we compared the piRNA expression profiles across a large number of samples (Fig. 4; online Supplemental Fig. 5). Raw sequencing data were analyzed, and across-data set normalization was conducted in the same way as for miRNAs. Importantly, since the number of piRNAs is relatively small, we carried out the normalization procedures by combining all data related to miRNAs and piRNAs to avoid potential bias due to the small number of variables. We observed that many piRNAs were highly specific to only a subset of data sets. For example, only 23% of piRNAs were detected with >1 read in $\geq 50\%$ of the samples included in online Supplemental Fig. 5, whereas 40% of miRNAs were found in a similar analysis. Due to this type of scarcity, we did not conduct a rank-based correlation analysis for piRNAs. Instead, the heat maps in Fig. 4 and online Supplemental Fig. 5 directly visualize piRNA expression levels with hierarchical clustering.

As shown in Fig. 4 and online Supplemental Fig. 5, CFS, embryonic stem (ES) cells, and skin cells were among those having the highest expression levels of piRNAs. The same observation can also be appreciated in online Supplemental Fig. 6, in which the expression levels of piRNAs are directly shown as empirical cumulative distributions. In addition, the data were examined for the fraction of piRNAs with at least 10 reads (after data normalization) in each sample (see online Supplemental Fig. 7). Again, CFS, ES cells, and skin cells were among those with the highest fraction of moderately or highly expressed piRNAs. In contrast, the piRNA expression in plasma (exosome-bound) was relatively low (see online Supplemental Figs. 5–7). These data suggest that that most salivary piRNAs may not have originated from circulating RNAs in blood. The similarity of piRNA profiles in CFS to those in ES cells and skin cells indicates that cells producing piRNAs in CFS (possibly salivary glands, oral mucosa, etc.) may have stem-like properties or regenerative capacity, which should be an area for further investigation.

IDENTIFICATION OF circRNAs IN HUMAN CFS

circRNAs constitute an emerging type of RNA recently highlighted in several studies involving different cell types and tissues *(20, 40–44)*. It is not yet known whether circRNAs exist extracellularly in body fluid. To examine this question, we constructed strand-specific sequencing libraries with a circRNA enrichment step using RNase R *(42)*. We developed a customized pipeline to identify unique back-spliced circular junctions within the sequencing reads (see online Supplemental Methods). It should be noted that this method searches for de novo circular junctions and does not depend on annotations of known exons and genes. Nevertheless, we observed that many of the predicted circRNAs were generated from known exons with canonical splice site signals (see online Supplemental Table 4). Interestingly, most such canonical circRNAs were also predicted as circRNAs in previous studies of intracellular RNA samples. A total of 95 putative canonical circRNAs were identified with at least 2 distinct circular junction reads among the 4 samples in this study or with 1 read but also reported as intracellular circRNA (http://circbase.org).

In addition to the canonical ones, many predicted circRNAs were not associated with canonical splice site signals (see online Supplemental Table 4). Because previous studies of intracellular RNA data did not consider such noncanonical circRNAs, we imposed a slightly more stringent criterion in calling such circRNAs by requiring at least 3 distinct reads overlapping the putative circular junction. A total of 327 noncanonical circRNAs were identified in this way. Thus, together, our study predicted 422 putative circRNAs in human CFS. Among these predictions, 28, 6, and 1 were common to at least 2, 3, and 4 individuals, respectively (see online Supplemental Table 4). The low degree of overlap may indicate that circRNAs are highly individual-specific. Alternatively, a much larger number of circRNAs may exist in each sample than detected in our study. Considering the large number of predicted intracellular circRNAs (>9000 in http://circbase.org), the latter possibility is likely true.

To date, the functional relevance of most intracellular circRNAs remains largely unknown. To gain some functional insights, we carried out a gene ontology analysis of the genes overlapping putative circRNAs in human CFS. Interestingly, a number of closely related categories were highly enriched, such as chemotaxis, inflammatory response, establishment of T cell polarity, cellular movement, actin cytoskeleton organization, and integrin-mediated signaling pathway (see online Supplemental Table 5). Overall, this result indicates that salivary circRNAs may be involved in intercellular signaling and inflammatory response. This observation is in line with the fact that inflammation is manifested via periodontal diseases, the most common disease in the oral cavity, and that exRNAs are now known to mediate cellular communications.

EXPERIMENTAL VALIDATION OF circRNAS IN HUMAN CFS

To validate the presence of circRNAs, primers capable of amplifying the predicted circular junctions (outward-facing primers, Fig. 5A) were used in a RT-PCR experiment (see online Supplemental Methods; online Supplemental Table 6). PCR product was visualized on a PAGE gel to confirm the expected size of the circular junction. Subsequently, DNA was purified and subject to Sanger sequencing. We randomly picked 6 putative circRNAs with the corresponding circles formed by 1 or 2 exons using canonical splice site signals. As shown in Fig. 5A, all 6 circular junctions were confirmed on the PAGE gel and, more importantly, with their exact sequences confirmed by Sanger sequencing. Thus, our data provide the first evidence that circRNAs exist in an extracellular body fluid.

Because a large number of our predicted circRNAs do not have canonical splice site signals, we conducted the above validation experiment on 3 additional noncanonical circRNAs. However, instead of direct Sanger sequencing, we conducted clonal sequencing of the TOPO-cloned PCR products because of the small size of these predicted circRNAs. The candidates were picked to represent 2 main categories of observed genomic locations that give rise to noncanonical circRNAs in our data: introns and long noncoding RNA (lncRNA) transcripts. Two of the 3 candidates were confirmed in this experiment (Fig. 5B). The circular junction (chr13:23270854_23270908) generated by a lncRNA was not confirmed, possibly owing to its low expression level and that only a limited number of clones were included for clonal sequencing. For the *EIF3E* (eukaryotic translation initiation factor 3, subunit E) intron–derived circRNA, the RNA-Seq reads led to discovery of multiple alternative circular junctions that differed by a small number of nucleotides (see online Supplemental Table 4). Our validation confirmed one of the most abundant forms. Interestingly, the ENCODE RNA-Seq data (polyA-fraction) derived from GM12878 cells (available at the UCSC Genome Browser, http://genome.ucsc.edu) included reads that correspond to this predicted circRNA, serving as an independent evidence to support its existence. The EIF3E intron harboring this circRNA is relatively long, and it is unlikely that the circular RNA was generated from the complete intron lariat after splicing. In contrast, the other confirmed circRNA (chr1:20979041_20979113) was derived from an intron of *DDOST* [dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit (non-catalytic)], the length of

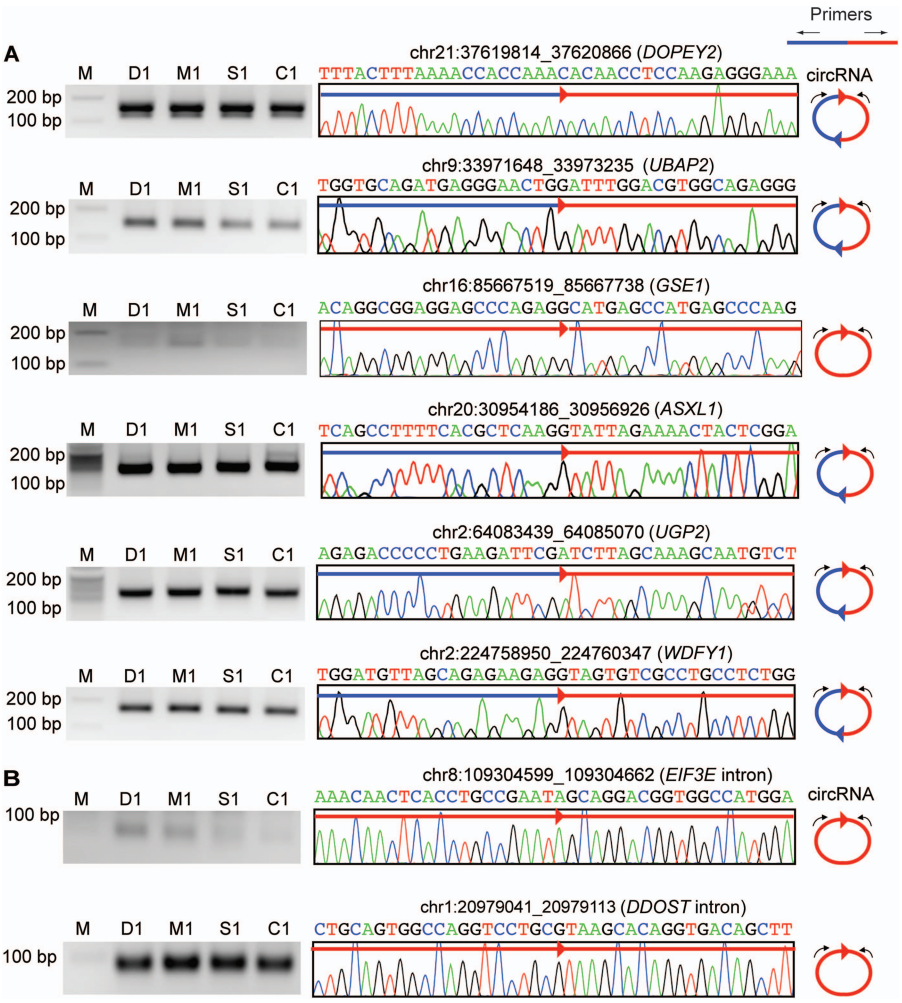

Fig. 5. circRNA expression in CFS. Experimental validation of circRNAs in CFS was carried out by RT-PCR.

Primers were designed to amplify the circular junctions, as illustrated. CFS samples were collected from the same subjects, but on different days, as for RNA-Seq. PAGE gel images are shown to visualize the presence of the PCR products. Representative Sanger sequencing traces are shown, together with the nucleotide sequences and genomic coordinates (hg19) of the circular junction. All Sanger-based sequences were identical to the predicted circular junctions via RNA-Seq. (A), Canonical circRNAs. Sanger sequencing was conducted on gel-purified PCR products. (B), Non-canonical circRNAs. Sanger sequencing was conducted on random clones amplified after TOPO cloning of the PCR products. *DOPEY2*, dopey family member 2; *UBAP2*, ubiquitin associated protein 2; *GSE1*, Gse1 coiled-coil protein; *ASXL1*, additional sex combs like transcriptional regulator 1; *UGP2*, UDP-glucose pyrophosphorylase 2; *WDFY1*, WD repeat and FYVE domain containing 1.

which is consistent with its generation from the intron lariat. Therefore, diverse biogenesis pathways may exist for these circRNAs, which need to be further investigated in the future.

In summary, we conducted a comprehensive study of the extracellular ncRNA profile of human saliva. To our best knowledge, this is the first report and validation of circRNAs in an extracellular body fluid. In addition, our study, for the first time, revealed the distinction and similarities between the landscapes of miRNAs and piRNAs in CFS and those of other body fluids and intracellular RNA samples. The validated presence of ncRNA in saliva provides novel insights into the biology and regulatory roles that saliva constituents can exert locally in the oral environment as well as systemically as saliva is being swallowed into the gastrointestinal tract. The insights generated in our work lay a foundation for future functional, mechanistic,

and translational discoveries related to ncRNAs in human saliva.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: D.T. Wong, RNAmeTRIX Inc.
Consultant or Advisory Role: D.T. Wong, PeriRxRNAmeTRIX Inc.
Stock Ownership: D.T. Wong, RNAmeTRIX Inc.; the University of California holds equity in RNAmeTRIX Inc.
Honoraria: None declared.
Research Funding: D.T. Wong, NIH grants UH2 TR000923-01 LC110207 from Department of Defense and 20PT-0032 from Tobacco-Related Disease Research Program; X. Xiao, NIH grants R01HG006264 and U01HG007013.
Expert Testimony: None declared.
Patents: D.T. Wong, intellectual property that D.T. Wong invented and was patented by the University of California has been licensed to RNAmeTRIX Inc.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We appreciate helpful discussions with Drs. David Chia and David Elashoff.

References

1. Wong DT. Salivaomics. J Am Dent Assoc 2012; 143:19S–24S.
2. Li Y, St John MA, Zhou X, Kim Y, Sinha U, Jordan RC, et al. Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res 2004;10: 8442–50.
3. Hu S, Wang J, Meijer J, Ieong S, Xie Y, Yu T, et al. Salivary proteomic and genomic biomarkers for primary Sjogren's syndrome. Arthritis Rheum 2007;56:3588–600.
4. Park NJ, Zhou X, Yu T, Brinkman BM, Zimmermann BG, Palanisamy V, Wong DT. Characterization of salivary RNA by cDNA library analysis. Arch Oral Biol 2007;52:30–5.
5. Park NJ, Li Y, Yu T, Brinkman BM, Wong DT. Characterization of RNA in saliva. Clin Chem 2006;52:988–94.
6. Walz A, Stuhler K, Wattenberg A, Hawranke E, Meyer HE, Schmalz G, et al. Proteome analysis of glandular parotid and submandibular-sublingual saliva in comparison to whole human saliva by two-dimensional gel electrophoresis. Proteomics 2006;6:1631–9.
7. Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park NH, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. Gastroenterology 2010;138:949–57;e1–7.
8. Spielmann N, Ilsley D, Gu J, Lea K, Brockman J, Heater S, et al. The human salivary RNA transcriptome revealed by massively parallel sequencing. Clin Chem 2012;58:1314–21.
9. St John MA, Li Y, Zhou X, Denny P, Ho CM, Montemagno C, et al. Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma. Arch Otolaryngol Head Neck Surg 2004;130:929–35.
10. Zhang L, Xiao H, Zhou H, Santiago S, Lee JM, Garon EB, et al. Development of transcriptomic biomarker signature in human saliva to detect lung cancer. Cell Mol Life Sci 2012;69:3341–50.
11. Lee YH, Kim JH, Zhou H, Kim BW, Wong DT. Salivary transcriptomic biomarkers for detection of ovarian cancer: for serous papillary adenocarcinoma. J Mol Med (Berl) 2012;90:427–34.
12. Zhang L, Xiao H, Karlan S, Zhou H, Gross J, Elashoff D, et al. Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. PLoS One 2010;5:e15573.
13. Tandon M, Gallo A, Jang SI, Illei GG, Alevizos I. Deep sequencing of short RNAs reveals novel micrornas in minor salivary glands of patients with Sjogren's syndrome. Oral Dis 2012;18:127–31.
14. Ogawa Y, Taketomi Y, Murakami M, Tsujimoto M, Yanoshita R. Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. Biol Pharm Bull 2013;36:66–75.
15. Burgos KL, Javaherian A, Bomprezzi R, Ghaffari L, Rhodes S, Courtright A, et al. Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. RNA 2013;19: 712–22.
16. Hu L, Wu C, Guo C, Li H, Xiong C. Identification of microRNAs predominately derived from testis and epididymis in human seminal plasma. Clin Biochem 2014;47:967–72.
17. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, Tuschl T. Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proc Natl Acad Sci U S A 2013;110:4255–60.
18. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, et al. Characterization of human plasma-derived exosomal rnas by deep sequencing. BMC Genomics 2013;14:319.
19. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. Nat Biotechnol 2014;32:453–61.
20. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. Nature 2013;495:384–8.
21. Lee YH, Zhou H, Reiss JK, Yan X, Zhang L, Chia D, Wong DT. Direct saliva transcriptome analysis. Clin Chem 2011;57:1295–302.
22. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.
23. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. Human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database (Oxford) 2010;2010:baq013. http://www.homd.org (accessed October 2014).
24. Gallo A, Tandon M, Alevizos I, Illei GG. The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. PLoS One 2012;7:e30679.
25. Li BS, Zhao YL, Guo G, Li W, Zhu ED, Luo X, et al. Plasma microRNAs, miR-223, miR-21 and miR-218, as novel potential biomarkers for gastric cancer detection. PLoS One 2012;7:e41629.
26. Xu J, Wu C, Che X, Wang L, Yu D, Zhang T, et al. Circulating micrornas, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. Mol Carcinog 2011;50:136–42.
27. Wang JF, Yu ML, Yu G, Bian JJ, Deng XM, Wan XJ, Zhu KM. Serum miR-146a and miR-223 as potential new biomarkers for sepsis. Biochem Biophys Res Commun 2010;394:184–8.
28. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 2008;18:997–1006.
29. Zhou X, Zhao F, Wang ZN, Song YX, Chang H, Chiang Y, Xu HM. Altered expression of miR-152 and miR-148a in ovarian cancer is related to cell proliferation. Oncol Rep 2012;27:447–54.
30. Farid WR, Pan Q, van der Meer AJ, de Ruiter PE, Ramakrishnaiah V, de Jonge J, et al. Hepatocyte-derived microRNAs as serum biomarkers of hepatic injury and rejection after liver transplantation. Liver Transpl 2012;18:290–7.
31. Kim SY, Jeon TY, Choi CI, Kim DH, Kim GH, Ryu DY, et al. Validation of circulating miRNA biomarkers for predicting lymph node metastasis in gastric cancer. J Mol Diagn 2013;15:661–9.
32. Thery C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol 2006;Chapter 3:Unit 3 22.
33. Sharma S, Rasool HI, Palanisamy V, Mathisen C, Schmidt M, Wong DT, Gimzewski JK. Structural-mechanical characterization of nanoparticle exosomes in human saliva, using correlative AFM, FESEM, and force spectroscopy. ACS Nano 2010; 4:1921–6.
34. Lau C, Kim Y, Chia D, Spielmann N, Eibl G, Elashoff D, et al. Role of pancreatic cancer-derived exosomes in salivary biomarker development. J Biol Chem 2013;288:26888–97.
35. Somel M, Guo S, Fu N, Yan Z, Hu HY, Xu Y, et al.

MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. Genome Res 2010;20:1207–18.
36. Joyce CE, Zhou X, Xia J, Ryan C, Thrash B, Menter A, et al. Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. Hum Mol Genet 2011;20: 4025–40.
37. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, et al. Regulation of microRNA expression and abundance during lymphopoiesis. Immunity 2010;32:828–39.
38. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010; 11:R106.
39. Sai Lakshmi S, Agrawal S. PiRNAbank: A web resource on classified and clustered piwi-interacting RNAs. Nucleic Acids Res 2008;36:D173–7.
40. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One 2012;7:e30733.
41. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 2013;495:333–8.
42. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with Alu repeats. RNA 2013;19:141–57.
43. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. PLoS Genet 2013;9:e1003777.
44. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, et al. Circular intronic long noncoding RNAs. Mol Cell 2013;51:792–806.